DOI:
10.1016/j.trit.2016.08.001

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

ScienceDirect

CAAI Transactions on Intelligence Technology xx (2016) 1−13
http://www.journals.elsevier.com/caai-transactions-on-intelligence-technology/

Original Article

Classification of epilepsy using computational intelligence techniques

Khurram I. Qazi [a,*], H.K. Lam [a], Bo Xiao [a], Gaoxiang Ouyang [b], Xunhe Yin [c]

[a] *Department of Informatics, King's College London, Strand, London, WC2R 2LS, United Kingdom*
[b] *State Key Laboratory of Cognitive Neuroscience and Learning, School of Brain and Cognitive Sciences, Beijing Normal University, No.19, XinJieKoWai St., HaiDian District, Beijing, 100875, PR China*
[c] *School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, PR China*

Available online ■ ■ ■

Abstract

This paper deals with a real-life application of epilepsy classification, where three phases of absence seizure, namely pre-seizure, seizure and seizure-free, are classified using real clinical data. Artificial neural network (ANN) and support vector machines (SVMs) combined with supervised learning algorithms, and *k*-means clustering (*k*-MC) combined with unsupervised techniques are employed to classify the three seizure phases. Different techniques to combine binary SVMs, namely One Vs One (OvO), One Vs All (OvA) and Binary Decision Tree (BDT), are employed for multiclass classification. Comparisons are performed with two traditional classification methods, namely, *k*-Nearest Neighbour (*k*-NN) and Naive Bayes classifier. It is concluded that SVM-based classifiers outperform the traditional ones in terms of recognition accuracy and robustness property when the original clinical data is distorted with noise. Furthermore, SVM-based classifier with OvO provides the highest recognition accuracy, whereas ANN-based classifier overtakes by demonstrating maximum accuracy in the presence of noise.

Keywords: Absence seizure; Discrete wavelet transform; Epilepsy classification; Feature extraction; *k*-means clustering; *k*-nearest neighbours; Naive Bayes; Neural networks; Support vector machines

1. Introduction

Epilepsy is a neurological condition such that it affects brain and the nervous system. It is a very commonly known neurological disorder and approximately 1% of general population is affected [1]. Only in the UK, around 1 in 100, more than half a million people suffer from epilepsy. There can be many causes of epilepsy and sometimes it is not possible to identify them. In the domain of epilepsy, seizure is referred to as an epileptic seizure and brain is the source. During an epileptic seizure normal functioning of the brain is disturbed for that certain time period, causing disruption on signalling mechanism between brain and other parts of the body. These seizures can put epilepsy patients at higher risk for injuries including fractures, falls, burns and submersion injuries, which are very common in children [2]. These injuries happen because seizure can happen anytime and anywhere without prior warning and the sufferer would continue his or her activity with an unconscious mind. If a system can effectively predict the pre-seizure phase (the transition time of the brain towards developing seizure), it could then generate an early warning alarm so that precautions can be taken by the sufferer.

Absence seizure is one from many forms of generalized epileptic seizures in which larger part of the brain is disturbed. These seizures are very short and sometime may go un-notice. The patient seems confused and may not remember the seizure afterwards. The complex spike-and-wave patterns generated by the brain during these seizures can be recorded on the electroencephalogram (EEG) and a neurologist can identify the three absence seizure phases namely seizure-free, pre-seizure and seizure [3,4]. To automate this process, EEG data is converted into a digital format and fed into a computerized

* Corresponding author.
E-mail addresses: khurram_ishtiaq.qazi@kcl.ac.uk (K.I. Qazi), hak-keung.lam@kcl.ac.uk (H.K. Lam), bo.xiao@kcl.ac.uk (B. Xiao), ouyang@bnu.edu.cn (G. Ouyang), xhyin@bjtu.edu.cn (X. Yin).
Peer review under responsibility of Chongqing University of Technology.

http://dx.doi.org/10.1016/j.trit.2016.08.001

seizure detection (classification) system, which can automatically recognize the input pattern. The two core modules of this classification system are: feature extraction and design of a classifier using these features. A feature extraction method extracts the most discriminative information from the EEG recordings, which means an ideal feature can have the property of differentiating among three phases of absence seizure.

In literature, there is a long list of methods that can be used to extract features from the EEG signals. These methods include Fourier Transforms (FT); good for analysing stationary signals, and Time Frequency Distribution (TFD); provides a tool for examining continuous segments of EEG signals. However, EEG signals are non-stationary in nature and conventional frequency analysis methods may not capture the full details of the brain signals. Lyapunov exponents; discussed the detection and prediction of epileptic seizure in Refs. [5], analysis of correlation structure [6], and high order spectral analysis (HOS) [7] are the examples of non-linear methods for EEG signal analysis in the domain of epilepsy. Advances in wavelet theory has also made it a very suitable for bio-medical signal processing. It has a built-in advantage of capturing repeated and irregular patterns. It can also deal with the analysis of transient and sudden signal changes [8,9]. This is possible because this technique provides variable window size, narrow at high and wide at low frequency levels. Furthermore [8], has discussed several different methods for EEG signal analysis and concluded that Wavelet Transform (WT) has more advantages over other methods. The next step in designing the classification system is to combine these extracted features with an appropriate learning method to design a classifier. These methods can be divided into traditional classification methods and modern learning algorithms also known as computationally intelligent algorithms or machine learning algorithms. Bayesian methods based on statistical theory, *k*-nearest neighbour (*k*-NN) and decision trees based on logical branching, are considered to be in the category of traditional classifiers. On the other hand, Support Vector Machine (SVM), Artificial Neural Network (ANN), *k*-Means Clustering (*k*-MC) and Self-organising Maps (SOMs) exhibit intelligent behaviour by learning the complex and non-linear patterns hidden inside the input feature vector, are considered to be computationally intelligent approaches. Below is the brief introduction about all the methods used in this research.

Classification systems based on Bayesian statistical theory have been widely and successfully used commercially [10]. This method gives a way to represent sensory evidence; features extracted from the raw data, and prior information about the problem in hand that is collected from domain knowledge. Considering the equal prior probabilities for all classes and hence, ignoring the hassle of obtaining the domain knowledge, the analysis becomes very straight forward. Although, it is a powerful and simple rule to handle and implement, yet estimating posterior probabilities from the data is a non-trivial task [11] and the distribution of data may not be uniform. *k*-NN rule as the name implies, classify an observation by giving it a label after probing the labels on all the *k*-NNs and making decision based on majority voting. Usually, Euclidean distance is used to measure the distances between the neighbouring instances. This algorithm mostly provides an acceptable performance in many applications [10] such as visual category recognition [12] and is also very easy to implement. However, *k*-NN algorithm suffers due to large memory requirements and also there is no logical way to choose the best *k* value, which would affect the classification problem and may not yield very good results.

Inspired by the human brain functioning and architecture, McCulloch and Pitts in 1940s presented a logical threshold unit (LTU) [13]. This basic idea of LTU has been generalized in many ways since then and it is the building block for modern ANNs. In 1962 a single layer perceptron neural network was introduced by Rosenblatt [14] by extending the idea of LTU along with a trainable network with adaptive elements [15]. Despite the huge success of this neural network model, it was only limited to solve linearly separable problems. A traditional feed-forward neural network [13] has three types of layers, i.e., input, hidden and output layers. Each layer consists of nodes connected in a layer-to-layer manner. The feed-forward neural networks can learn any smooth non-linear functions in a compact domain to any degree of accuracy and are considered to be universal approximators. They have many applications such as classification, prediction, clustering and approximation [13]. The drawback of this technique is that it requires a lot of parameters to be tuned for training the neural network. Also, there is no set of rules for finding the number of hidden layers and neurons in these layers. Furthermore, ordinary neural networks suffer from the "over-fitting" and "local optimization" problems [16].

The SVM method maps the non-linear and inseparable data from an input space into a higher dimensional feature space where the data would then be linearly separable [17]. This task is accomplished by utilising the concept of separating hyper planes [17]. Instead of computing a mapping function, the use of kernel function saves the computational demand especially for feature mapping function of higher dimensional space. The SVM algorithm aims to maximize the margin (the region separating the support vectors on either side of the hyperplane) and tries to find an optimal hyperplane. Hence, also called the maximal margin classifier. Although, the training parameters for SVM technique are very few, it can still be a computationally time consuming and highly complex. Nonetheless, SVM has a good generalization ability, solution to the over-fitting problem and also performs well in a high dimensional feature space.

k-MC is an unsupervized learning technique that clusters samples based on similar properties. Whereas, *k* represents the number of clusters or classes. Within a cluster the samples are similar but different from the samples grouped in other clusters. This is an iterative process in which samples are grouped together according to their closet mean and cluster positions are adjusted until these positions do not change for some iterations. The advantage of this technique is that it does not require labels with the input examples. The convergence is also faster if *k* is small. However, if clusters are of different

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

sizes, densities and not globular then the results are poor. The results may also vary depending on the initial location of the cluster centres.

From the above discussion, we can conclude that the advantage of using machine learning approaches over the traditional classification methods is that the classification system does not need to know much about the input data in the beginning but the characteristics of the input data are learnt by the learning algorithm. Once the classification system has been built, replacing a learning algorithm with any other would not require any changes in the system. Considering this advantageous factor, classifiers such as ANN and SVM have been widely used in designing the epilepsy detection systems [3,9,6,18]. Also, features that are highly discriminative not only yields better recognition accuracy but also speeds up the process. Recently [3], has shown a very good recognition accuracy of seizure phases by extracting features in the time domain as well as in the frequency domain using FT.

In this paper, we first present a different strategy for extracting features from the EEG signals. In the time domain local features, based on windowing filtered mechanism, and global features, based on the entire signal length, are calculated. The signal is then decomposed into different levels using Discrete Wavelet Transform (DWT) and features at each level are calculated. In the next phase, we compare different machine learning algorithms in terms of their recognition accuracy based on the extracted features. Traditional classification methods provide an acceptable performance in many applications and can be used as a benchmark for other techniques [10]. Comparisons have also been performed between both types of classification methods, traditional and intelligent algorithms. The robustness property has also been tested using noise-contaminated data.

The organisation of this paper is as follows. Section 2 explains the mechanism of digital data acquisition from epilepsy patients and also discusses the process of feature extraction method for designing the epilepsy classification system. It then explains how the three computational intelligence techniques; SVM, ANN and *k*-MC work. Section 3 presents classification results that include design of classifiers using SVM, ANN and *k*-MC and evaluation of these classifiers against two traditional classification methods, NBC and *k*-NN. The paper ends with concluding remarks in Section 4.

2. Epilepsy classification system

Fig. 1 shows the design cycle of epilepsy classification system used in this research. This is an iterative process in which the system is updated based on the output produced by the learning algorithms. The details of this design scheme are discussed in the following sections.

2.1. Data acquisition

Using 10-20 international standard electrode placement system, with all 19 electrodes (FP1, FP2, F3, F4, C3, C4 P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz) the data has been collected from the epilepsy patients suffering from the absence seizure. These recordings were taken in the Peking University People's Hospital, China, and the patients (6 males and 4 females), aged 8−21 years old, have agreed to use this data for research and publishing results [3]. To record the neural activity, EEG data was samples at 256 Hz, filtered from 0.5 to 35 Hz bands using Neurofile NT digital video EEG system [3]. Extracted data has been divided into three phases of absence seizure; seizure-free, pre-seizure and seizure, by an epilepsy neurologist. These phases are separated based on the criteria of 1) the interval between the seizure-free phase and beginning point of seizures is greater than 15 s, 2) the interval is 0 and 2 s prior to seizure onset, and 3) the interval is the first 2 s of the absence seizure [3]. Each dataset has 110 samples and each sample size is 19×512. Fig. 2 shows the example of 19-channel EEG recordings of these three seizure phases. From the figure, seizure-free and seizure phases are quite obvious and easier to classify. However, the brain shifts from seizure-free to an absence seizure (pre-seizure phase) looks almost the same and is hard to distinguish. The generalized spike-wave discharges with a repetition rate of 3 Hz are typically associated with clinical absence seizures. More details about the data collection can be located in Refs. [19−21].

2.2. Feature extraction

The purpose of this stage is to eliminate the redundancy in the EEG signals by selecting the discriminative features from the raw data. The process also helps in reducing the size of the input feature vector. In the case of EEG recordings, we have 19 channels from the 19 EEG electrodes (sensors) as shown in Fig. 2. The research in absence epilepsy has demonstrated that not all 19 channels are of the same importance. Successful experiments in Ref. [3] has also established that only 10 electrodes (FP1, FP2, F3, F4, C3, C4, F7, F8, T3 and T4) out of total 19 have the properties that can help in classifying the three seizure phases. A feature vector was then formed based on these 10 selected channels.

To capture the brain functioning over a period of time, local and global values of the same features were calculated from the EEG signals in the time domain. Global values were calculated based on the entire signal length and local features were extracted using a windowing function. The importance of the features in the time domain is evident from Fig. 2, in which there are no sudden or abrupt changes in the brain recordings during the seizure-free and pre-seizure phases. Such activities are very well captured by the time domain analysis. Energy (E), range (R), standard deviation (SD), sum of absolute values (SAV), mean absolute values (MAV) and variance (Var) of each selected channel were extracted in the time domain to form the part of the feature vector.

DWT decomposes the signal into several levels and each level represents a particular coarseness of the signal. At each level, the signal is passed through the high pass filter (HPF), which acts as the mother wavelet, and the low pass filter (LPF) that acts as a mirror version of the corresponding HPF. The output of each level is the downsized signal by a factor of 2.

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

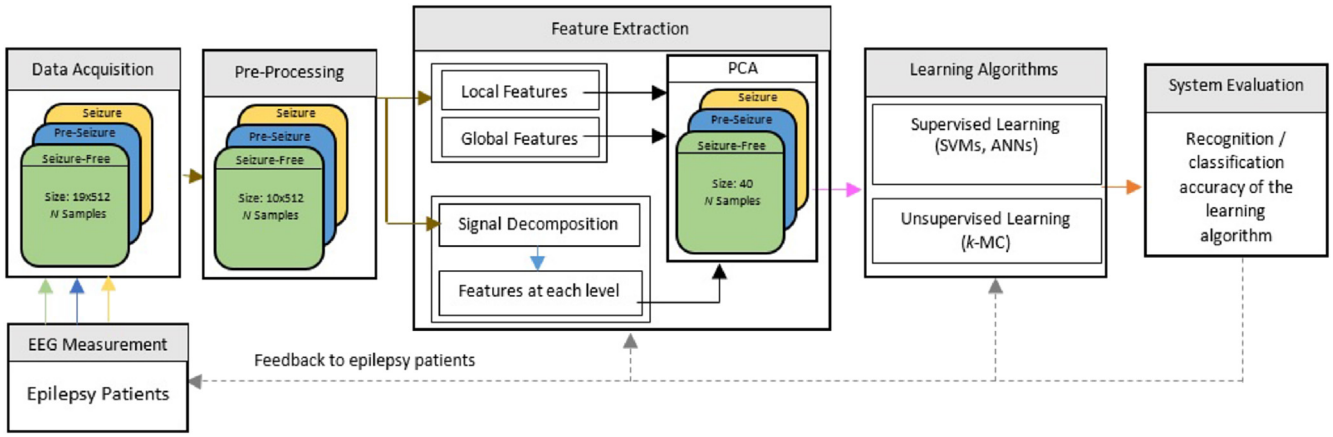


Fig. 1. Design cycle of epilepsy classification system.

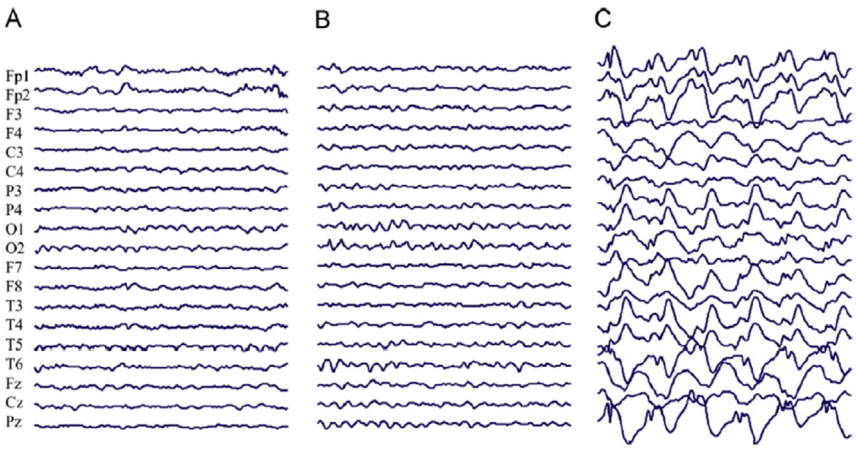


Fig. 2. EEG recordings. Seizure-Free (A), Pre-Seizure (B) and Seizure (C) phases.

Approximation (A1), consists of low frequency components, and detail (D1), consists of high frequency components, are the coefficients produced by LPF and HPF respectively. For the next level of decomposition, A1 is further decomposed and the whole process is repeated until the desired level of decomposition is achieved. For one dimensional signal $x(t)$ the continuous wavelet transform (CWT) is defined by Morlet-Grossman [22]:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^{*}\frac{(t-b)}{a}dt \tag{1}$$

where ψ is the analysing wavelet, a is the scalar parameter, and b is the position parameter. The CWT is converted into discrete wavelet by using binary scale factors. Hence, from (1) [23]:

$$W(j,b) = \frac{1}{\sqrt{2j}} \int_{-\infty}^{\infty} x(t)\psi^{*}\frac{(t-b)}{a}dt \tag{2}$$

Which means A1 and D1 are the first split of signal $x(t)$. Then from using A1 the next split is A2 and D2 and the process continues until the desired level is reached. Fig. 3 depicts the DWT decomposition scheme of a signal of length n and $\downarrow^2$ shows that the signal is downsized by a factor of 2 at each level.

We have used Daubechies D4 wavelets, which is most widely used DWT for EEG signal analysis. D4 uses four scaling and wavelet coefficients [9]. The scaling coefficients (LPF) are:

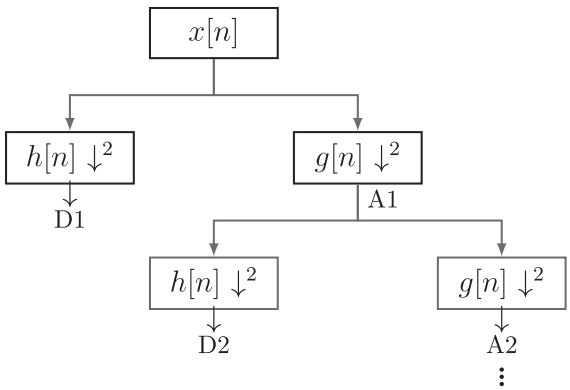


Fig. 3. DWT decomposition scheme.

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, \quad h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

And the wavelet coefficients (HPF) are:

$$g_0 = h_3, \quad g_1 = -h_2, \quad g_2 = h_1, \quad g_3 = -h_0$$

EEG signal has been decomposed into 5-levels and at each level energy, range and standard deviation has been calculated. In addition to these features, waveform length (WL), average amplitude change (AAC) and difference absolute standard deviation (DASD) [24] have also been computed at each level of decomposition to form the rest of the feature vector. Combining all the features from time domain as well as from the decomposed signal resulted in a large feature vector, which is considered to be computationally expensive in the classification process and can also generate poor recognition accuracy. Fig. 4 depicts the whole process of feature extraction discussed above. As from Section 2.1, the individual electrode provides 512 brain readings (values) and only 10 channels are selected to build the feature vector. Hence, form each electrode, 192 local values using a window size of 16 and 6 global feature values are extracted. The values of individual electrode is then decomposed to 5-levels. $D_j \cdots D_L$ represent the detail values and A_L represents approximate values as shown in Fig. 3. Where $j = 1 \cdots L$ and $L = 5$. Features are extracted at each level giving 36 values per electrode. Principal component analysis (PCA) has been used to reduce the dimension (d) of the feature vector. A 40 dimensional feature vector is decided to be used as an input to design the classifier and to perform the classification experiments.

2.3. Computational intelligence techniques

Intelligent techniques take feature vector as an input and use a learning strategy these algorithms learn the patterns in the features vector. After the learning (training) process has finished any new data sample is classified based on the trained classifier. This phase is called classification phase. According to the learning procedure, these techniques can be divided into supervised and unsupervised learning. In supervised learning, the algorithm takes example inputs (training dataset) and their corresponding output (class labels). It then learns a general rule of mapping inputs to outputs. SVM and ANN follow this strategy. In an unsupervised learning mechanism, no output is attached with the input examples. The data is grouped and clustered based on their natural or similar characteristics. k-MC is an unsupervised learning algorithm. In this research, we have explored SVM, ANN and k-MC.

2.3.1. Support vector machine

A binary class classification problem $y^{(i)} \in [+1, -1]$ with a training set $[\mathbf{X}^{(i)}]_{i=1}^{N} \in \Re^d$, and the corresponding outputs (class labels) $[y^{(i)}]_{i=1}^{N}$ can be stated as:

$$\mathbf{W}^T \mathbf{X}^{(i)} + w_0 \geq 1 \quad \forall_i : y^{(i)} = 1 \tag{3}$$

$$\mathbf{W}^T \mathbf{X}^{(i)} + w_0 \leq -1 \quad \forall_i : y^{(i)} = -1 \tag{4}$$

Where $\mathbf{W} \in \Re^d$ is the weight vector and w_0 is the bias. However, not all the classification problems are linearly separable. Such problems are solved using the soft-margin method and the classifier is known as soft-margin classifier. Maximising the margin and minimising the number of misclassified points, we can formulate the constraint optimization problem for the soft-margin method as:

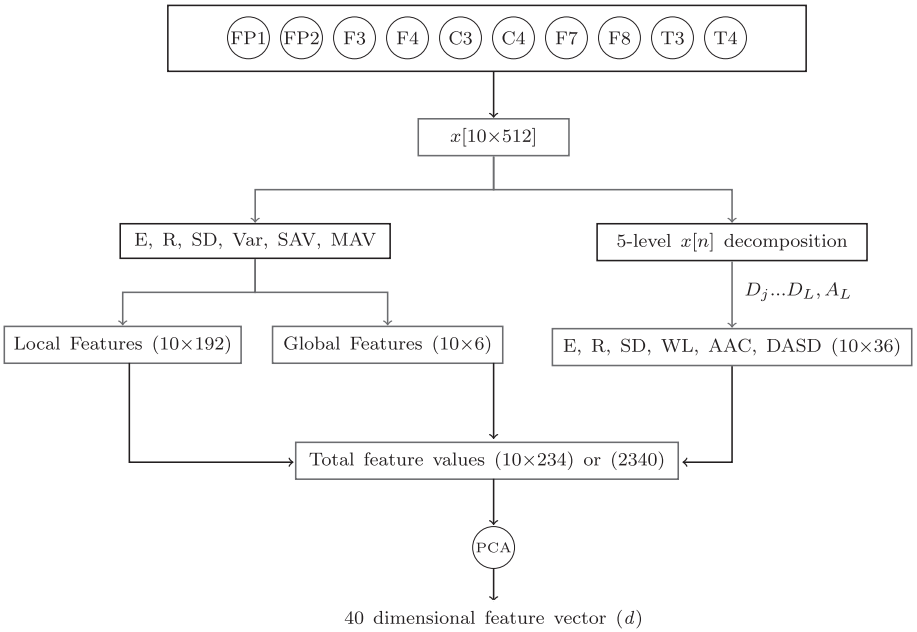


Fig. 4. Feature extraction process.

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

$$\min_{\mathbf{W},w_0,\zeta} J(\mathbf{W},\zeta) = \frac{1}{2}||\mathbf{W}||^2 + C\sum_{i}^{N}\zeta^{(i)}$$
$$subject\ to: y^{(i)}\left(\mathbf{W}^T\mathbf{X}^{(i)} + w_0\right) \geq 1 - \zeta^{(i)} \quad (5)$$
$$\zeta^{(i)} \geq 0 \quad and \quad i = 1,2,....,N$$

Where ζ is a slack variable, which represents the upper bound training error and $C > 0$ trades off margin size and training error. Replacing C with 0 in (5) becomes the linearly separable case. Non-linear Support Vector Classifiers (SVCs) use the kernel trick, which maps the original feature space to another higher dimensional space and through this mapping process the classification task may become easier. In this study we use the following three kernel functions [25].

$$Linear: K\left(\mathbf{X}^{(i)}, \mathbf{X}^{(i)'}\right) = \sum_{j=1}^{d} x_j^{(i)} x_j^{(i)'} \quad (6)$$

$$Ploynomial: K\left(\mathbf{X}^{(i)}, \mathbf{X}^{(i)'}\right) = \left(1 + \sum_{j=1}^{d} x_j^{(i)} x_j^{(i)'}\right)^q \quad (7)$$

$$RadialBasis: K\left(\mathbf{X}^{(i)}, \mathbf{X}^{(i)'}\right) = \exp\left(-\gamma \sum_{j=1}^{d}\left(x_j^{(i)} x_j^{(i)'}\right)^2\right) \quad (8)$$

where $K(\bullet,\bullet)$ represents a kernel function, d is the dimension of the feature vector, q is the degree of the polynomial and γ sets the influence boundary for the radial basis kernel function. SVM is a binary classifier that can only handle two-class classification problems. To overcome this shortcoming different strategies are used to combine the binary SVM classifiers to tackle multiclass classification problems. As in our case, we are required to classify three phases of absence seizure. We have compared three different methods to combine binary SVM classifiers, which are Binary Decision Tree (BDT), One Vs One (OvO), and One Vs All (OvA).

BDT method requires $C-1$ classifiers for C-classes. Converting a three-class classification problem into a two-class classification problem, SVM1 is trained on samples from seizure phase (class 3) with labels, -1, and samples from seizure-free and pre-seizure phases (classes 1&2) combined together as one class with labels, $+1$. SVM2 is then trained on samples from seizure-free phase (class 1) with labels, $+1$, and pre-seizure phase (class 2) samples with labels, -1. A test sample $\mathbf{X}^{(*)}$ is processed through both the SVMs and to obtain the category of this new sample the results of both SVMs are then combined as shown in Fig. 5. The classification accuracy depends on the SVMs in the upper levels.

$\frac{C(C-1)}{2}$ binary SVM classifiers are required for C-classes using OvO method. Thus, for three-class problem, it requires 3 binary SVMs. SVM1 is trained on samples from class 1 (seizure-free) and class 2 (pre-seizure), SVM2 on samples from class 2 and class 3 (seizure), and SVM3 on samples from class 3 and class 1. For a new observation (sample) $\mathbf{X}^{(*)}$, it chooses final class label with majority votes. This whole process is summarized in Fig. 6.

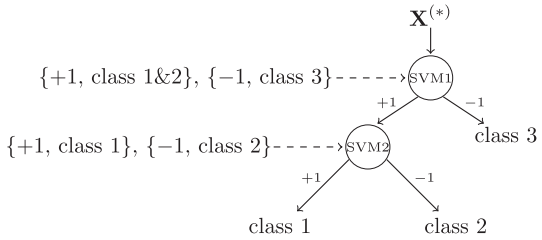


Fig. 5. Three-class SVM-BDT.

C-binary SVM classifiers are needed for C-class problem when using OvA method. Therefore, 3 binary classifiers for the problem in hand. Three-class problem is again divided into two-class classification problem and the training procedure is shown in Fig. 7, where class 1 is seizure-free, class 2 is pre-seizure and class 3 is seizure phase. A new sample $\mathbf{X}^{(*)}$ is passed through all three SVMs and the SVM having highest value will decide the final class label.

2.3.2. Artificial Neural Networks

A fully-connected three-layer feed-forward neural network (FFNN) is shown in Fig. 8. Each layer is connected to its previous and the next layer. These connections are associated with connection weights. Hidden layers also contain an array of artificial neurons, called processing units and an activation function, $f(\bullet)$, for each neuron. A general feed-forward operation for classification purposes can be written as [16]:

$$g_k\left(\mathbf{X}^{(i)}\right) \equiv z_k^{(i)} = f\left(\sum_{p=1}^{n_H} w_{kp} f\left(\sum_{j=1}^{d} w_{pj} x_j^{(i)} + w_{p0}\right) + w_{k0}\right) \quad (9)$$

where the subscript j indexes units in the input layer and p indexes units in the hidden layer. w_{pj} denotes the input-to-hidden layer weights at the hidden unit p. The subscript k indexes in the output layer and n_H denotes the number of hidden units. d represents the dimension of the ith input sample. The advantage of using FFNN compared with SVM is that for C-classes (outputs), the network can learn C discriminant functions z_k and a training input $\mathbf{X}^{(i)}$ is assigned to the function $z_k^{(i)}$ that has the maximum value from all other functions. Usually, these functions are learnt through a well-known backpropagation algorithm. The goal of this algorithm is to find a set of weight values for all the connections, such that it can minimize the error between the actual and desired output. A new sample $\mathbf{X}^{(*)}$ is classified using the trained FFNN.

To achieve the maximum recognition accuracy of the three seizure phases, we have tried different number of hidden layers, hidden nodes and various combination of activation functions in this study. The best recognition accuracy was achieved by using 4 hidden-layer FFNN. The first hidden layer has 22 neurons and uses radial basis activation function. The second hidden layer has an array of 8 neurons and each node uses tangent sigmoid activation function. The third and fourth hidden layers contain 10 and 4 neurons respectively, and both

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

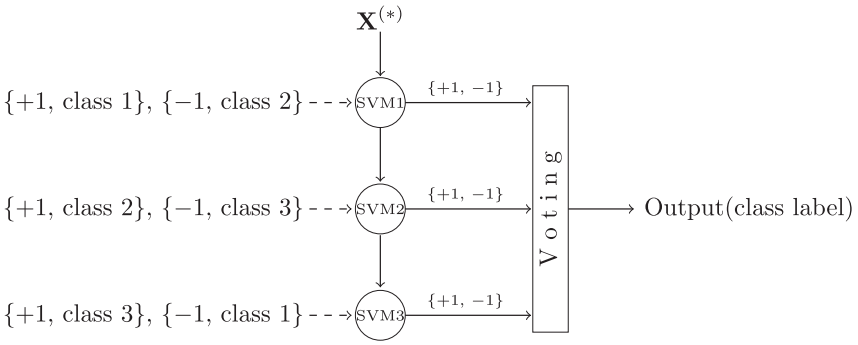


Fig. 6. Three-class SVM-OvO.

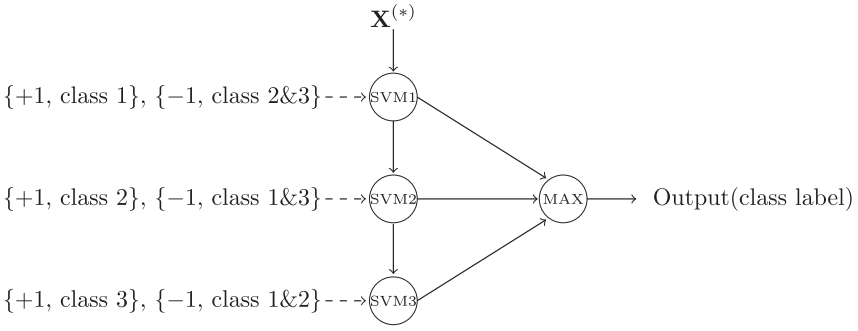


Fig. 7. Three-class SVM-OvA.

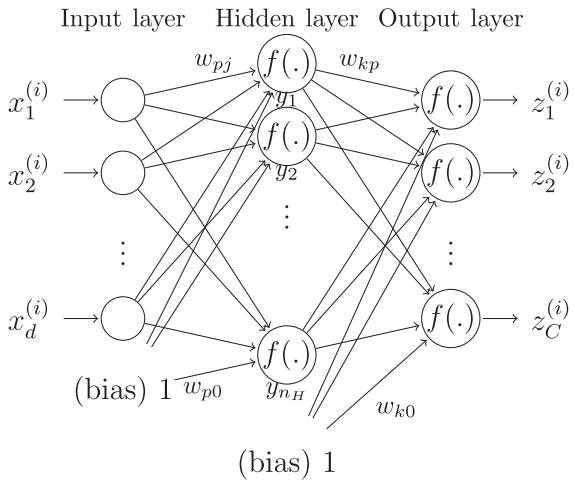


Fig. 8. Three-layer feed-forward fully connected neural network.

layers are using logarithmic tangent sigmoid transfer function. The output layer is consisting of three neurons. Both input and output layers are using linear activation function. The FFNN takes the samples of size d as an input along with the corresponding class labels and then produces the output of size 3. The class labelling scheme has been shown in Table 1. The training parameters and procedure for this 4 hidden-layer FFNN architecture have been provided in the results section.

2.3.3. k-means clustering

The problem of clustering can be defined as the sum of square Euclidean distances of each cluster from its mean $\mathbf{m}_k$:

Table 1
Class labelling scheme for FFNN.

Class name	Class labels
Seizure-free	1, 0, 0
Pre-seizure	0, 1, 0
Seizure	0, 0, 1

$$J_e = \sum_{k=1}^{C} \sum_{\mathbf{X}^{(i)} \in D_k} \left|\left| \mathbf{X}^{(i)} - \mathbf{m}_k \right|\right|^2 \tag{10}$$

This criterion defines clusters as their mean vectors $\mathbf{m}_k$ in the sense that it minimizes the sum of the squared lengths of the error, $\mathbf{X}^{(i)} - \mathbf{m}_k$. Where C represents the total number of clusters (classes), $\mathbf{X}^{(i)}$ is the ith sample, and D_k is the kth cluster. Hence, a partition is optimal if it minimizes J_e and also called minimum variance partition. In our case, the data was clustered into three groups (seizure-free, pre-seizure and seizure phases) and a new sample was assigned to the cluster that has minimum distance from these three clusters.

It is worth mentioning that after the training process, all the parameters and weights of the classifiers are fixed values and can be used in the real-life applications. The computational cost for the classifiers to give out the results are all very low and can be ignored in most of the cases.

3. Classification results

From Section 2.1, each dataset (seizure-free, pre-seizure and seizure) has 110 samples. To design classifiers using extracted features from the EEG signals as discussed in

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

Section 2.2, and to perform the classification experiments, each dataset is divided into 22 testing samples and 88 training samples. The training dataset is used to train the classifier. After the training process, the classifier is tested using the testing dataset to evaluate its recognition accuracy for any new pattern (sample). The percentage of recognition accuracy for both training and testing datasets are calculated separately based on correctly classified samples within each individual dataset.

3.1. Classifiers

To perform the experiments with SVM-based classifiers, a soft-margin classifier has been used in all the multiclass SVM classifiers. When polynomial kernel function (7) is used, we choose $q = 5$. When radial basis kernel function (8) is used, we choose $\gamma = 0.2$. These values are used in all experiments which are achieved by trial and error approach for the best performance.

The FFNN architecture discussed in Section 2.3.2 was trained with training samples along with their corresponding class labels for 30 epochs and with initial learning rate of 10^{-4} using Levenberg-Marquardt backpropagation algorithm. Mean squared error criteria was used to calculate the error between the outputs produced by the network and the actual class labels. For each successful step, the learning rate was decreased by 10^{-1} and for every failure, it was incremented by 5. The network takes 40 dimensional feature vector as an input, and produces the output of size 3. $(1, 0, 0)$ represents an observation from class seizure-free, $(0, 1, 0)$ corresponds to a sample from pre-seizure phase and $(0, 0, 1)$ classifies a seizure phase.

For k-MC, 100 iterations were performed to determine the cluster centres. However, clustering is a random process, which starts with random cluster centres. During these iterations, if the clusters did not change their positions, new clusters centres were determined to avoid the local minima problem. This process replicated itself 5 times.

3.2. Evaluation

To evaluate the success of our feature extraction scheme and to equate the recognition accuracy between learning algorithms and traditional classification methods, we have also selected to perform experiments with naive Bayes and k-nn classification methods. To estimate the data distribution for naive Bayes classifier, we have employed two ways to estimate the input data distribution. The one uses the normal Gaussian function (Naive Bayes with Gaussian function NBN) and the other uses the kernel smoothing function (Naive Bayes with kernel smoothing function NBK). k-NN has been tested for four different values of k (i.e., 1, 3, 5 and 7). Class labels 1, 2, and 3 have been used for seizure-free, pre-seizure and seizure samples respectively, for both the traditional classification methods.

The recognition accuracy (classification performance) of all the methods are evaluated. Firstly, the experiments were performed using the original dataset, without any noise. In the EEG process the noise however is inevitable due to certain uncertainties in the measurement process, such as, signal conversion, filtering, amplifying and environment conditions. To reflect these factors present in the real-life patient's clinical data, how the underlying seizure recognition system would behave in the presence of noise and also to investigate the robustness property of each algorithm, the data has been contaminated with different levels (0.05, 0.1, 0.2, 0.5 and 1.0) of random noise.

The recognition accuracy of all the methods are presented in Tables 2–7, where SVM-OvA represents SVM classifier with one vs all strategy, SVM-OvO is SVM classifier with one vs one method and SVM-BDT shows the results obtained using SVM classifier with binary decision tree method. Linear, polynomial and radial basis kernel functions show the results against each individual kernel method discussed in Section 2.3.1. NN is the neural network classifier consisting of four hidden layer architecture, NBC is naive Bayes classification method, which uses two different density estimation methods, normal (NBN) and kernel smoothing function (NBK). k-NN is shown with different values of k (i.e., 1, 3, 5 and 7). And finally, k-MC shows the k-means clustering results.

In Table 2, the recognition accuracy (%) of both training and testing datasets is shown where "Average" represents the average accuracy for all three classes (seizure phases). It can be seen that the highest average testing accuracy is 93.9394% which is achieved by SVM-OvO using polynomial kernel function and the corresponding average training accuracy is 99.6970%. Whereas, 1-NN has obtained the best average training recognition accuracy of 100%. On the other hand, the lowest average testing accuracy is 66.6667% which is achieved by k-MC. Talking about successfully predicting the pre-seizure phase, SVM-OvO with polynomial kernel function is again leading with 90.9091% testing recognition accuracy. Also, traditional classification methods are not able to predict the pre-seizure phase with good recognition accuracy. NBC with kernel function has improved the average testing accuracy from 69.1358% to 72.8395%. However, NBC provides a poor performance in predicting pre-seizure phase.

The fact that the real-life data will not be smooth and it will contain noise due to certain factors mentioned above. Tables 3–7 summarizes the results of testing data classification when original testing dataset is contaminated with different levels of noise. Addition of noise is a random process. Thus to collect the statistical information, each experiment was repeated 10-times on the test dataset only. These experiments were performed without training the classifier on the noisy dataset. "Worst", "average", "best" and "std" (standard deviation) show overall values for all the three seizure phases. The worst individual is the lowest individual accuracy obtained after these experiments.

Generally, the recognition accuracy is decreasing when the noise level is increasing. The recognition accuracy of SVM with polynomial kernel function has badly degraded in all the multiclass SVM methods. However, SVM with linear kernel function has shown good robustness property in the presence of different noise levels. Among all the classification methods,

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

Table 2
Summary of training and testing recognition accuracy for EEG signals with original datasets.

Classifier	Recognition accuracy (%). Noise level = 0							
	Training				Testing			
	Average	Seizure-free	Pre-seizure	Seizure	Average	Seizure-free	Pre-seizure	Seizure
SVM-OvA								
Linear	93.0303	92.0455	87.0455	100	86.0606	86.3636	77.2727	94.5455
Polynomial	99.6970	100	99.0909	100	80.9091	90.9091	88.1818	63.6364
RBF	93.8636	91.1364	90.4546	100	86.9697	86.3636	74.5454	100
SVM-OvO								
Linear	94.4697	92.7273	90.6818	100	85.4546	80.9091	76.3636	99.0909
Polynomial	99.697	100	99.0909	100	93.9394	91.8182	90.9091	99.0909
RBF	93.7879	90.9091	90.4546	100	87.2727	86.3636	75.4546	100
SVM-BDT								
Linear	94.3939	92.9546	90.2273	100	86.9697	85.4546	75.4546	100
Polynomial	99.7727	100	99.3182	100	91.5152	86.3636	89.0909	99.0909
RBF	93.8636	90.9091	90.6818	100	86.6667	83.6364	76.3636	100
NN								
	97.5806	97.5903	95.1219	100	92.6829	96.2963	82.1428	100
k-MC								
	70.2811	62.6506	48.1928	100	66.6667	62.9630	40.7407	96.2963
NBC								
NBN	71.8876	90.3615	28.9157	96.3855	69.1358	85.1852	22.2222	100
NBK	85.9438	100	57.8313	100	72.8395	88.8889	33.3333	96.2963
k-NN								
1-NN	100	100	100	100	85.1852	85.1852	74.0741	96.2963
3-NN	93.1727	93.9759	85.5422	100	80.2469	81.4815	62.9630	96.2963
5-NN	91.9679	90.3614	85.5422	100	80.2469	81.4815	62.9630	96.2963
7-NN	91.5663	89.1566	85.5422	100	87.6543	88.8889	77.7778	96.2963

Table 3
Summary of testing recognition accuracy for EEG signals under testing dataset subject to noise level of 0.05.

Classifier	Recognition accuracy (%). Noise level = 0.05						
	Overall (seizure-free, pre-seizure, seizure)				Worst individual		
	Worst	Average	Best	Std	Seizure-free	Pre-seizure	Seizure
SVM-OvA							
Linear	83.0303	84	86.3636	1.3787	82.7273	70	91.8181
Polynomial	77.2727	79.9394	82.4242	1.8729	87.2727	83.6364	60
RBF	85.1515	86.0606	86.6667	0.7107	83.636	70	100
SVM-OvO							
Linear	83.0303	85.3333	86.0606	1.2963	77.2727	72.7273	98.1818
Polynomial	88.7879	91.0909	93.3333	1.6353	85.4546	81.8182	99.0909
RBF	84.8445	85.8788	87.2727	0.9486	84.5454	70	100
SVM-BDT							
Linear	84.8485	85.9394	87.8788	1.1459	81.8182	72.7273	99.0909
Polynomial	89.697	90.9091	92.4242	0.982	89.0909	72.2727	96.3636
RBF	85.7576	86.0606	86.6667	0.3711	84.5455	70	100
NN							
	87.8049	91.3009	95.1219	1.9932	88.8889	78.5714	92.5925
k-MC							
	65.4321	65.8025	66.6667	0.5964	62.9630	40.7407	92.5926
NBC							
NBN	69.1358	69.1358	69.1358	0	85.1852	22.2222	100
NBK	72.8395	72.8395	72.8395	0	88.8889	33.3333	96.2963
k-NN							
1-NN	85.1852	85.1852	85.1852	1×10^{-14}	85.1852	74.0741	96.2963
3-NN	79.0124	80	80.2469	0.5200	77.7778	62.9630	96.2963
5-NN	80.2469	80.7407	82.7161	1.0400	81.4815	62.9630	96.2963
7-NN	85.1852	86.5432	87.6543	1.0800	81.4815	74.0741	96.2963

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

Table 4
Summary of testing recognition accuracy for EEG signals under testing dataset subject to noise level of 0.1.

Classifier	Recognition accuracy (%). Noise level = 0.1						
	Overall (seizure-free, pre-seizure, seizure)				Worst individual		
	Worst	Average	Best	Std	Seizure-free	Pre-seizure	Seizure
SVM-OvA							
Linear	80.9091	82.0606	83.333	1.0584	81.8182	67.2727	88.1818
Polynomial	78.7879	79.8788	80.9091	0.9959	89.0909	82.7273	60.9091
RBF	85.4546	86	86.9697	0.6211	83.6364	71.8182	100
SVM-OvO							
Linear	83.6364	84.8485	78.7576	0.8835	76.3636	73.6364	98.1818
Polynomial	89.697	90.9091	92.1212	1.0276	87.2727	83.6364	97.2727
RBF	84.8485	85.6364	86.3636	0.5907	83.6364	70.9091	100
SVM-BDT							
Linear	82.4242	84.2424	86.3636	1.469	78.1818	66.3636	99.0909
Polynomial	89.697	90.8485	92.1212	1.1009	86.3636	83.6364	97.2727
RBF	84.5455	85.8788	86.9697	0.9241	84.5455	69.0909	100
NN							
	87.8048	90.0406	93.9024	2.0527	88.8889	71.4286	92.5926
k-MC							
	64.1975	65.5556	66.6667	0.7008	62.9630	37.037	92.5926
NBC							
NBN	69.1358	39.1358	69.1358	0	85.1852	22.2222	100
NBK	72.8395	72.8395	72.8395	0	88.8889	33.3333	96.2963
k-NN							
1-NN	85.1852	85.1852	85.1852	1×10^{-14}	85.1852	74.0741	96.2963
3-NN	79.0124	80.1235	80.2469	0.3900	77.7778	62.9623	96.2963
5-NN	80.2469	81.4815	82.7161	1.3000	81.4815	62.9630	96.2963
7-NN	83.9506	85.5556	87.6573	1.3000	77.7778	74.0741	96.2963

Table 5
Summary of testing recognition accuracy for EEG signals under testing dataset subject to noise level of 0.2.

Classifier	Recognition accuracy (%). Noise level = 0.2						
	Overall (seizure-free, pre-seizure, seizure)				Worst individual		
	Worst	Average	Best	Std	Seizure-free	Pre-seizure	Seizure
SVM-OvA							
Linear	81.5152	82	82.7273	0.5505	78.1818	70.9091	88.1818
Polynomial	76.6667	78.6061	79.697	1.1853	86.3636	80	60
RBF	84.5455	85.8788	86.3636	0.7902	85.4546	68.1818	100
SVM-OvO							
Linear	83.6364	84.424	85.1515	0.5907	78.1818	71.8182	99.0909
Polynomial	90	90.7273	91.8182	0.7902	84.5455	80.9091	98.1818
RBF	85.1515	85.5152	86.0606	0.3951	82.7273	70	100
SVM-BDT							
Linear	84.2424	84.7872	85.4546	0.4494	80	70.9091	100
Polynomial	89.697	90.9091	92.1212	1.0276	88.1818	80.9091	97.2727
RBF	83.6364	85.697	87.2727	1.7595	80	62.2727	100
NN							
	85.3659	88.6179	91.4634	1.4792	85.1860	67.8571	92.5926
k-MC							
	64.1975	65.1852	66.6667	0.7808	62.9630	37.037	92.5926
NBC							
NBN	69.1358	69.1358	69.1358	0	85.1852	22.2222	100
NBK	71.6049	72.5926	72.8395	0.521	85.1852	33.3333	96.2963
k-NN							
1-NN	85.1852	85.1852	85.1852	1×10^{-14}	85.1852	74.0741	96.2963
3-NN	79.0124	80.1235	80.2469	0.3900	77.7778	62.9630	96.2963
5-NN	80.2469	82.2222	83.9506	1.5616	77.7778	62.9630	96.2963
7-NN	83.9506	85.0617	86.4198	0.9110	77.7778	74.0741	96.2963

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

Table 6
Summary of testing recognition accuracy for EEG signals under testing dataset subject to noise level of 0.5.

Classifier	Recognition accuracy (%). Noise level = 0.5						
	Overall (seizure-free, pre-seizure, seizure)				Worst individual		
	Worst	Average	Best	Std	Seizure-free	Pre-seizure	Seizure
SVM-OvA							
Linear	80.9091	81.9394	82.7273	0.7295	80.9091	64.5455	90.9091
Polynomial	70.303	73.6364	77.2727	2.9458	90	59.0909	52.7273
RBF	82.4242	83.0303	83.3333	0.4286	84.5454	61.8182	100
SVM-OvO							
Linear	83.3333	84.1818	85.7576	1.0141	78.1818	69.0909	99.0909
Polynomial	82.4242	83.9294	86.3636	1.8433	87.2427	59.0909	97.2727
RBF	82.1212	83.5758	85.1515	1.1216	82.7273	69.6364	100
SVM-BDT							
Linear	82.4242	84.2424	86.3636	1.469	78.1818	66.3636	99.0909
Polynomial	83.0303	86.1212	90.3030	3.1302	89.0909	59.0909	98.1818
RBF	83.3333	84.4242	86.0606	1.2604	86.336	61.8183	100
NN							
	85.3659	87.3577	90.2439	1.4844	88.8889	64.2857	92.5926
k-MC							
	62.9630	64.5679	65.4321	0.8333	62.9630	37.037	88.8889
NBC							
NBN	69.1358	69.1358	69.1358	0	85.1852	22.2222	100
NBK	71.6049	72.5926	72.8395	0.521	85.1852	33.3333	96.2963
k-NN							
1-NN	85.1852	86.2963	87.6543	0.7000	85.1852	70.3704	96.2963
3-NN	77.7778	80	82.7161	1.5176	77.7778	59.2593	96.2963
5-NN	77.7778	79.5062	82.7161	1.8600	77.7778	59.2593	96.2963
7-NN	77.7778	80.4939	82.7161	1.4000	81.4815	55.5556	96.2963

Table 7
Summary of testing recognition accuracy for EEG signals under testing dataset subject to noise level of 1.

Classifier	Recognition accuracy (%). Noise level = 1						
	Overall (seizure-free, pre-seizure, seizure)				Worst individual		
	Worst	Average	Best	Std	Seizure-free	Pre-seizure	Seizure
SVM-OvA							
Linear	78.4849	80.9091	83.3333	2.0885	79.0909	68.1818	86.3636
Polynomial	67.8788	71.5738	75.5455	2.7641	88.1818	57.2727	54.5454
RBF	81.5152	82.8485	84.5455	1.1658	85.4546	59.0909	100
SVM-OvO							
Linear	81.2121	83.2121	85.1515	1.3985	77.2727	67.2727	98.1818
Polynomial	80.9091	82.3636	83.6364	1.2928	89.0909	55.4546	97.2727
RBF	81.8182	82.9697	84.5455	1.0365	84.5455	60	100
SVM-BDT							
Linear	82.1212	83.2427	83.9394	0.7235	76.3636	69.0909	99.0909
Polynomial	81.2121	82.7273	84.5455	1.469	86.3636	55.4546	97.2727
RBF	80.6061	82.5455	84.2424	1.3985	83.0303	58.1818	84.2424
NN							
	80.4878	85.8943	89.0244	2.0677	88.8889	53.5714	88.8889
k-MC							
	62.9630	64.4444	65.432	0.7808	62.9630	37.037	88.8889
NBC							
NBN	69.1358	70.2469	70.2469	0.39	85.1852	22.2222	69.1358
NBK	70.3704	71.1111	71.6049	0.638	81.4815	33.3333	92.5926
k-NN							
1-NN	83.9506	85.5556	87.6543	1.0200	85.1852	66.6667	96.2963
3-NN	77.7778	79.2593	80.2469	0.9740	77.7778	59.2593	96.2963
5-NN	77.7778	80.1235	83.9506	1.8813	77.7778	55.5556	96.2963
7-NN	79.0124	80.4939	81.4815	0.9740	81.4815	59.2593	96.2963

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

the NN classifier has shown the maximum recognition accuracy of 85.8943% at highest noise level of 1. From the discussion and the results presented, it is fair to conclude that NN has a good generalization ability which is not very much degraded under the increasing levels of noise.

4. Conclusion

This paper has presented different supervised (SVM and NN) and unsupervised (*k*-MC) learning algorithms for classifying epilepsy seizure phases. Computationally intelligent techniques NN and SVM have proved to be very good in recognising and classifying the complex and complicated patterns in the input data (EEG signals). The performance of these algorithms has been compared with two traditional classification methods, NBC and *k*-NN. It can be concluded that NN and SVM have demonstrated the best recognition accuracy compared with traditional classification methods and an unsupervised learning algorithm, *k*-MC. Furthermore, NN and SVM both showed robustness property by maintaining best results even when the input data is contaminated with different noise levels. In addition, the feature extraction method is also introduced in the paper, which is able to gain rich information of the signal at a moderate dimension of the feature vector. Also, the NN method outperforms the SVM methods for some of the noisy dataset cases. In future, more classification methods based on neural networks such as self-organising maps and deep learning architecture will be explored and try to find the method, which is best at both noisy case and noise free case.

Acknowledgment

The work described in this paper was partly supported by King's College London, China Scholar Council, National Natural Science Foundation of China (No. 61172022) and Foreign Experts Scheme of China (No. GDW20151100010).

References

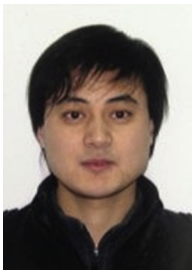
[1] P. Kwan, T.J. Brodie, Neurother 6 (2006) 397–406.
[2] E.C. Wirrell, Epilepsia 47 (1) (2006) 79–86.
[3] H.K. Lam, Udeme Ekong, Bo Xiao, Gaoxiang Ouyang, Hongbin Liu, K.Y. Chan, Sai Ho Ling, Neurocomputing 149 (Part C) (2015) 1177–1187.
[4] U. Ekong, H.K. Lam, B. Xiao, G. Ouyang, H. Liu, K.Y. Chan, S.H. Ling, Neurocomputing 199 (2016) 66–76.
[5] S. Osowski, B. Swiderski, A. Cichocki, A. Rysz, COMPEL Int. J. Comput. Math. Electr. Electron. Eng. Emerald 26 (5) (2007) 1276–1287.
[6] James R. Williamson, Daniel W. Bliss, and David W. Browne, Epileptic seizure prediction using the spatiotemporal correlation structure of intracranial EEG, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2011) pp. 665–668.
[7] K.C. Chua, V. Chandran, U.R. Acharya, C.M. Lim, J. Med. Syst. 35 (6) (2011) 1563–1571. Springer.
[8] Amjad S. Al-Fahoum, Ausilah A. Al-Fraihat, ISRN Neurosci. 2014 (2014) 7. Hindawi Publishing Corporation, Article ID 730218.
[9] Ibrahim Omerhodzic, Samir Avdakovic, Amir Nuhanovic, Kemal Dizdarevic, Kresimir Rotim, Energy distribution of EEG signal components by wavelet transform, in: Dumitru Baleanu (Ed.), Wavelet transforms and their recent applications in Biology and Geoscience, 2012, p. 310. InTech ISBN: 978-953-51-0212-0.
[10] Anil K. Jain, Robert P.W. Dulin, Jianchang Mao, IEEE Tran Pattern Anal Mach. Intell. 22 (1) (2000) 4–37.
[11] G.P. Zhang, IEEE Trans. Syst. Man Cybern. Soc. Appl. Rev. 30 (4) (2000) 451–462.
[12] H. Zhang, A.C. Berg, M. Maire, J. Malik, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2 (2006) 2126–2136.
[13] Anil K. Jain, Jianchang Mao, K.M. Mohiuddin, IEEE Comput. Sci. Eng. 29 (3) (1996) 31–44.
[14] G.L. Marcialis, F. Roli, Pattern Recognit. Lett. 26 (12) (2005) 1830–1839.
[15] F. Ham, I. Kostanic, Principles of Neurocomputing for Science and Engineering, McGraw-Hill, New York, 2001.
[16] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, second ed, John Wiley and Sons Inc, U.K, 2006, p. 680. ISBN: 978-81-265-116-7.
[17] S. Yaman, J. Pelecanos, W. Zhu, Unifying PLDA and polynomial kernel SVMS, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2013, pp. 7698–7701.
[18] Abdulhamit Subasi, Ahmet Alkan, Etem Koklukaya, M. Kemal Kiymik, Neural Netw. 18 (7) (2005) 985–997.
[19] M. Thulasidas, C. Guan, J. Wu, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (1) (2006) 24–29.
[20] T.C. Ferree, P. Luu, G.S. Russell, D.M. Tucker, Clin. Neurophysiol. 112 (3) (2001) 536–544.
[21] O'Keefe, M.L. Recce, Hippocampus 3 (3) (1993) 317–330.
[22] S. Sanei, J.A. Chambers, EEG Signal Processing, John Wiley and Sons Ltd, 2007, p. 312. ISBN: 978-0-470-02581-9.
[23] X. Zhang, W. Diao, Z. Cheng, Wavelet Transform and Singular Value Decomposition of EEG Signal for Pattern Recognition of Complicated Hand Activities, Springer, 2007, pp. 294–303 vol. 4561 of the series lecture notes in computer science.
[24] Angkoon Phinyomark, Pornchai Phukpattaranont, Chusak Limsakul, Expert Syst. Appl. 39 (2012) 7420–7431.
[25] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning with Application in R, Springer Science and Business Media, New York, 2013, p. 440. ISBN: 978-1-4614-7138-7.

Bo Xiao received bachelor and master (Hons.) degree from Chongqing University, China, in 2010 and 2013, respectively. He is currently working towards his Ph.D. degree in King's College London. His current research interests include computational intelligence, T-S fuzzy model based fuzzy control, polynomial fuzzy model based fuzzy control and interval type-2 fuzzy logic.

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001

H.K. Lam received the B.Eng. (Hons.) and Ph.D. degrees from the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, in 1995 and 2000, respectively. During the period of 2000 and 2005, he worked with the Department of Electronic and Information Engineering at The Hong Kong Polytechnic University as Post-Doctoral Fellow and Research Fellow respectively. He joined as a Lecturer at King's College London in 2005 and currently a Reader. His current research interests include intelligent control systems and computational intelligence.

Xunhe Yin received the bachelor and doctoral degree from Harbin Science and Technology University and Harbin Institute of Technology, in 1989 and 2000, respectively. Currently, he works in Beijing Jiaotong University, Beijing, China as Professor. His current research interests include Information and control technology of intelligent transportation system, Control theory and control engineering.



Gaoxiang Ouyang received the B.S. degree in automation and the M.S. degree in control theory and control engineering both from the Yanshan University, Hebei, China, in 2002 and 2004, respectively, and the Ph.D. degree from the Department of Manufacturing Engineering, City University of Hong Kong , in 2010. He currently serves as an Associate Professor in the School of Brain and Cognitive Sciences, Beijing Normal University, Beijing, China. His research interests include biosignal analysis, neural engineering, and dynamics system.

Please cite this article in press as: K.I. Qazi, et al., Classification of epilepsy using computational intelligence techniques, CAAI Transactions on Intelligence Technology (2016), http://dx.doi.org/10.1016/j.trit.2016.08.001